

Individuation through infrastructure: Get Full Text Research, data extraction and the academic publishing oligopoly

Moore, S.

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Moore, S 2020, 'Individuation through infrastructure: Get Full Text Research, data extraction and the academic publishing oligopoly', *Journal of Documentation*, vol. (In-press), pp. (In-press).

<https://dx.doi.org/10.1108/JD-06-2020-0090>

DOI 10.1108/JD-06-2020-0090

ISSN 0022-0418

Publisher: Emerald

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Individuation through infrastructure: *Get Full Text Research*, data extraction and the academic publishing oligopoly

Abstract

This article explores the recent turn within academic publishing towards ‘seamless access’, an approach to content provision that ensures users do not have to continually authenticate in order to access journal content. Through a critical exploration of *Get Full Text Research*, a service developed collaboratively by five of the world’s largest academic publishers to provide such seamless access to academic research, the article shows how publishers are seeking to control the ways in which readers access publications in order to trace, control and ultimately monetise user interactions on their platforms. Theorised as a process of individuation through infrastructure, the article reveals how publishers are attempting an ontological shift to position the individual, quantifiable researcher, rather than the published content, at the centre of the scholarly communication universe. The implications of the shift towards individuation are revealed as part of a broader trend in scholarly communication infrastructure towards data extraction, mirroring a trend within digital capitalism more generally.

Introduction

In 2019, five of the world’s largest academic publishers announced the development of *Get Full Text Research* (GetFTR), a service designed to indicate to readers whether they have permission to access a scholarly article and then directly provide access if they do (Get FTR, n.d.). When launched, the service will be a user-facing graphical interface to indicate whether

or not a reader has access to an article. It is designed to sit within the digital infrastructures of services that allow it, aiming to operate in the background and thus reduce the inconvenience of continual authentication and ‘password fatigue’ on the behalf of readers (Sanchez et al., 2016). The purported benefit of GetFTR to the reader is seamlessness: rather than being required to continually authenticate their credentials between publishers and across platforms, readers may simply log into the GetFTR service once in order to have their institutional accounts recognised with participating publishers. This will remove the friction involved when users transition between indexes, databases, journals and other platform providers for scholarly communication.

GetFTR itself is designed and led solely by five of the largest commercial publishers, American Chemical Society, Elsevier, Springer Nature, Taylor & Francis Group and Wiley. These five publishers have been previously described as the academic publishing ‘oligopoly’ for their collective market share of over half of all academic articles published around the world (Larivière et al., 2015b). Academic publishing is a concentrated industry with significant market power exercised by a handful of small actors, many of whom enjoy annual profit margins of over 30% (Buranyi, 2017; Larivière et al., 2015a). In the face of piracy, open access and academic social networks such, GetFTR represents an attempt by publishers to control the ways in which researchers access content, thus resulting in further consolidation and control of academic publishing by a small number of publishers.

As an infrastructural intervention, GetFTR raises important questions about the ways in which a digital infrastructure may order the behaviour of users. Specifically, why are publishers interested in controlling how their content is accessed, and what are the resulting trade-offs between user convenience and control over one’s information consuming

behaviour? Many of these trade-offs are based on privacy: in return for the convenience of continuous authentication, publishers are able to collect data on users in order to further monetise their behaviour through product design and other services. Through a critical exploration of the relationship between infrastructure and user behaviour, this article explores how GetFTR is representative of a recent strategic shift in the business models of large commercial publishers from the mere selling of content to a more diversified approach based on extracting data analytics from the users on their platforms. While this move towards data extractivism is certainly not a new phenomenon, as others have noted (Couldry & Mejias, 2019; Morozov, 2018; Zuboff, 2019), it does reveal something pertinent to academic publishing infrastructure design, namely a strategy by commercial publishers to create user-centric rather than content-centric modes of access in order to track behaviour and measure performance. The implications of this shift, in which universities are complicit, are both profound and pernicious, as this article explores.

Background

The ways in which researchers access journal articles is rapidly changing. Having transitioned almost entirely from print to online in the past two decades, research access is now shifting further from subscription access to open access. It is estimated that by 2025, 44% of all articles ever published will be open access and that 70% of all journal views will be of open access articles (Piwowar et al., 2019). Yet this trajectory towards openness will not be immediate, and it is likely that research access will be a mixed model of open and subscription content for many years to come. During this mixed economy, readers will continue to navigate the many and various subscriptions and databases provided through academic libraries, alongside content found on the open web.

For academics within universities, journal subscriptions are negotiated by their libraries and provided to the reader through a variety of databases and portals. There is no single way of finding and accessing research; each researcher's approach is different and highly personalised, usually entailing a variety of web searches and database queries (Fried Foster, 2014). Once a subscription-access resource is identified, researchers may then seek to determine whether they have access to the article in question through their library's subscriptions. This is achieved through a range of mechanisms, such as IP authentication on campus and VPN access off campus, or through single sign-on authentication, such as Shibboleth or OpenAthens, that require a user's institutional username and password (Arsenault et al., 2020).

From a user experience perspective, research access is a frustrating exercise, as explored in a report by the *Research Access for the 21st Century* (RA21) initiative:

Today's researchers face a confusing diversity of options to facilitate remote access to scholarly information resources, such as VPN servers, proxy servers, various access code and registration schemes, and third-party software solutions. These difficulties in navigating today's remote access solutions impede research, frustrate users, and may encourage fully entitled users to resort to illicit, pirate websites. (NISO, 2019, p. 12).

RA21 was a multi-stakeholder initiative backed by the STM publishing industry group that represents the academic publishing industry. RA21 had a goal of 'facilitating a simple user experience for users of scholarly information resources' (NISO, 2019). Ostensibly, the initiative positioned itself as responding to the frustrations documented above: research access is cumbersome in the face of multiple passwords, repeated sign-ins and the lack of a

single standard way to access digital resources. However, these frustrations had long persisted with web-based research access and the true catalyst for the push for improved authentication is the publishing industry's loss of control over how researchers access content.

With the rise of open access, academic social networks and piracy websites, publishers are no longer in total control over how academics access published content. Open access publications may proliferate across the web through the use of Creative Commons licenses, allowing permission-free sharing of content without requiring mediation by publisher portals (Eve, 2014, p. Chapter 3). Open access articles appear in a range of locations, including faculty websites, institutional/subject repositories, and academic social networks, all of which legally circumvent publisher portals. Many of these articles are syndicated through browser plugins such as Unpaywall and the Open Access Button that find and provide direct access to content behind paywalls where an open version is available. Furthermore, academic social networking websites have been particularly successful at diverting traffic away from publisher websites, with ResearchGate a notable example. ResearchGate hosts content uploaded by users and scraped from various sources on the web, resulting in a corpus of '130 million' downloadable research papers according to ResearchGate (n.d.). ResearchGate itself is also a considerable source of articles discovered by Google Scholar, with 10% of articles in the database linking directly to the ResearchGate site (Jamali & Nabavi, 2015).

However, many of the papers uploaded to academic social networks are available without the copyright holder's consent. Publishers frequently maintain copyright of the articles they publish, meaning that authors are only able to share them according to publisher restrictions. Yet many authors share their papers irrespective of these restrictions. For example, one study

used a sample to estimate that roughly 50% of non-open access articles available through ResearchGate were uploaded without the copyright holder's permission (Jamali, 2017). This situation has led large commercial publishers, notably those responsible for GetFTR, to issue take-down notices to ResearchGate requiring them to remove content that infringes copyright (Van Noorden, 2017). In addition to ResearchGate, piracy websites such as Sci-Hub also redirect readers away from legally accessible copies from publishers. One study estimated that nearly all published scholarly articles in existence are available through Sci-Hub (Himmelstein et al., 2018), while another study highlighted the many millions of users downloading articles from Sci-Hub from the vast majority of countries around the world (Bohannon, 2016).

Either through legitimate or illicit means, researchers are able to download scholarly articles from a range of locations. This means that publishers are no longer the sole providers of the content they publish and that a plurality of content sources have emerged. Should this trend continue, as it likely will as publishing rapidly moves open access, publishers will lose their monopolistic ability to monetise both the content *and* the metadata generated by user interactions with that content. For example, publishers currently monetise usage statistics to show how frequently accessed their articles are, a practice that informs negotiations over the price of journal subscriptions and 'Big Deal' packages of journal bundles (Schonfeld, 2019b). These statistics will be lower where readers access content in ways publishers cannot track, thus reducing their negotiation power. Publisher negotiation power is all the more vital at a period in which universities are cancelling Big Deal subscriptions in response to publisher pricing strategies and feet-dragging over affordable open access options (McKenzie, 2019). The Scholarly Publishing and Academic Resources Coalition maintains a list of universities

that have cancelled various Big Deal packages, showing how widespread the practice is (SPARC, n.d.).

But more generally, publishers are losing control of the *users* within their ecosystems. As open access content proliferates, and university/funder policies require authors to share articles via repositories, publishers are unable to track user interactions with content in order to monetise transactional analytics and tailor services and content to users. User tracking is an additional revenue stream that requires maximum interaction with the services that publishers provide, from submission through to publication and beyond. The ability to track users by rendering them traceable, trackable individuals, I argue, is informing the infrastructural strategies of the academic publishing industry and is one of the key motivations for *Get Full Text Research*.

***Get Full Text Research* as infrastructural intervention**

In response to the loss of control resulting from different research access pathways, publishers have developed *Get Full Text Research* (GetFTR) to encourage researchers to return to publisher platforms. When a reader discovers an article from an index or discovery service that has adopted GetFTR, the reader will be presented with a graphical interface that shows whether they have access to the article, either through subscriptions or open access, and they will be taken directly to the resource if they do. The process is described on the GetFTR website using the example of a hypothetical researcher named Eva:

If Eva has recently authenticated at her institution she'll go straight to the article. If she hasn't, she will be directed to her university login page to check her credentials.

No need to stop first at the publisher website, no decisions to make about which login

to use. The publisher seamlessly delivers the article to her there and then, without her having to know different login details or publisher website addresses. And if she doesn't have access? The publisher can then provide an alternative version of the content, such as a preprint or a read-only version. This all happens instantly, regardless of where Eva is working, whether at home, in the university library, the lab, her local coffee shop or anywhere else. Eva hasn't had to download anything or install additional software. (Get FTR, n.d.)

GetFTR is presented as simple and seamless; it will remove the friction inherent in accessing content between different publishers and will thus provide a better user experience for research access, as argued by the publishing technologist Ian Mulvaney (2019). Moreover, if the reader does not have access through their institutional subscriptions, then GetFTR suggests that an 'alternative version' would be made available through a preprint or read-only version. This means that publishers will retain traffic from open access and non-legal sources through the provision of research beyond the version of record (e.g., a preprint) or that may have limited functionality (e.g., it is not downloadable) but must be accessed on the publisher's platform through its article viewer software. In either case, publishers retain control of the pathways to access, including the usage data, reading behaviour and statistics based on them.

GetFTR is an infrastructural intervention that will allow publishers to retain control of access pathways by intervening in the digital infrastructures that order researcher behaviour. It does this through the usage of a technology called 'Seamless Access' that allows single sign-on authentication across different publisher platforms. Based on the recommendations of the RA21 project described above, Seamless Access is a service that authenticates the user's

browser and prevents them from requiring to continually authenticate their details each time they need access to a publisher's resources. GetFTR deploys this technology to indicate to a user whether they have access to content, providing direct access if so or offering alternative pathways if not, thus extending the 'seamlessness' beyond the content to which a user subscribes and towards alternate versions where available.

It is worth at this stage thinking about the relationship between seamlessness and infrastructure, particularly how infrastructure has the ability to blend in the background while impacting on the behaviour of those interacting with it. For Leigh Star, infrastructure is itself *defined* in part by seamlessness, i.e., it only becomes 'visible' when it breaks down, stops working and its seams are revealed (Star, 1999, p. 382). The invisibility of infrastructure also allows it to embody certain standards that may not be immediately apparent to the end user (ibid). Infrastructures are also modular and interconnected, meaning that adapting or re-designing them is not a simple process and instead requires negotiations between a range of different actors and user groups. This is to say that although infrastructure may appear benign and natural, it actually has a profound ability to influence and shape user behaviour, as the designers of GetFTR understand.

Understood according to Star's work, scholarly communication infrastructures are actually quite visible, rather than seamless, and reveal themselves in a manner of ways through a user's experience. For example, the connection between publisher portals and academic databases is not a smooth process. The seams are revealed in every literature search and continual need to authenticate across modular platforms. Each authentication represents a breakdown in the user's experience and makes visible the mediating points between various infrastructures at hand, even for those with access to all the resources they need. In fact, the

continual need to authenticate serves as a reminder of certain economic and institutional privileges associated with research access: research access is anything but seamless for those without access to a well-stocked university library. The connecting points between search engines and published resources can be navigated by those with access, whereas they are insurmountable for those without. Either way, GetFTR seeks to erase the visible liminality between digital infrastructures by removing this need to continually authenticate. Through the push for ‘seamlessness’, it thus hopes to make research access *more infrastructural*, according to Star’s definition, in the sense that it is less visible to the end user. Though this process may be more convenient for the user, it also leaves them open to being ordered and tracked within the publisher-controlled ecosystem. As users move seamlessly between digital infrastructures, their behaviours and pathways are shaped by the standards that each infrastructure ‘embodies’ (to use Star’s terminology discussed above).

Though GetFTR is an infrastructural intervention that appears relatively minor, it is actually illustrative of a broader shift by the major commercial publishers towards user individuation. By making the authentication process easier (and less visible), and discouraging users from seeking alternative forms of access, publishers hope to keep users interacting with their products and services, not only those that provide access to journal content but also those from across the knowledge production ecosystem. It is a strategy that will allow publishers to track and monetise users from submission to publication by capitalising on the interaction data that takes place on their platforms. As the research consultant Roger Schonfeld writes in the *Scholarly Kitchen* blog:

Publishers will naturally elect to route not only entitled users to their sites for subscription access but also all users to their sites for gold open access publications

— a huge opportunity for them. And it seems likely that most will provide an alternative access option to those who do not have subscription access, seeing it as a chance to gather user data and explore an array of sales opportunities (Schonfeld, 2019a).

This ‘huge opportunity’ is not one that requires the sale of journal content or subscriptions, but rather entails tracing and monetising the interactions of users to the greatest extent possible. In order to do this, users must be individuated, which is to say interpellated by publishers as separate, trackable entities.

In the current academic publishing system, readers primarily interact with publisher services to access the content they require by actively authenticating via their institutional subscriptions. In the GetFTR system, the user’s institutional subscriptions are continually authenticated and so their entitlements are known in advance. The current system is content-centric while GetFTR is user-centric, i.e., the former approach confirms whether or not the journal content is available to the reader, while the latter attempts to connect the user to a version of the content based on their access privileges. Schonfeld describes this as a process of ‘recentering product offerings around researchers’ in a way that focuses products on user ‘identity management’ rather than on the content being published. For Schonfeld, this turn to identity management is ‘absolutely vital to enable research collaboration, assessment, and showcasing, along with product personalization and an array of adjacent opportunities’ (Schonfeld, 2019a). GetFTR is illustrative of this strategy to individuate readers into single, quantifiable, ‘personalizable’ units through its promise of convenient, seamless access.

Many librarians and technologists have reacted to GetFTR (and seamless access more generally) with caution over the privacy issues of such a service. For Peter Murray, GetFTR has the potential to share ‘personally identifiable’ information such as a patron’s IP address (Murray, 2019). Similarly, Lisa Hinchliffe points to the privacy issues associated with the Seamless Access technology that under pins GetFTR, writing: ‘The predecessor project, RA21 raised many concerns related to control over and privacy of user data and the future of publisher support for proxy and IP based authentication, access pathways that are valued and broadly implemented in academic libraries’ (Hinchliffe, 2019). These concerns are also outlined by Hinchliffe in article in *Science*: ‘[P]ublishers employing the service might eventually require universities to provide details about users who log in, such as their names and academic departments, as a condition for accessing journal articles. They could potentially use those data for business purposes, perhaps selling it to third-party data aggregators’ (Brainard, 2019). However, for Ian Mulvany, user privacy is less of a concern because ‘GetFTR explicitly states that they don’t get any information about the user’ (Mulvany, 2019).

But the concerns about privacy and data extraction reveal the direction of travel rather than publishers’ actual ability to track users under GetFTR. The developers of Open Access Button, a service for automatically requesting copies of paywalled journal articles from their authors, argue that ‘[u]ser privacy is under threat. But, as commercial players lock up this space, we should expect the first policies here to be ‘not as bad as expected’ as they push for mass adoption. However, policies change, and we should be clear about who’s in the driver’s seat as they do’ (Open Access Button, 2019). Once GetFTR becomes a standard way of accessing journal content – by both subscribing and non-subscribing readers – one can expect privacy rules to be relaxed for maximum user data extraction. This is why GetFTR is

indicative of the strategy by the publishing oligopoly to individuate and subsequently track users across their ecosystems. The approach is especially noteworthy for how publishers are collaborating rather than competing on ways to individuate and track users, illustrating that extractivism is the next frontier for the business strategies of commercial publishing more generally, but that this can only be achieved by firstly making users trackable on publisher platforms.

The collaboration around GetFTR is therefore indicative of a more general strategy to nurture the conditions for data extractivism, or to create what Julie E. Cohen terms a *biopolitical public domain*: ‘a repository of raw materials that are there for the taking and that are framed as inputs to particular types of productive activity’ (Cohen, 2018). The biopolitical public domain presupposes that interactions on the web are ‘available and potentially valuable’ and ‘raw’, which is to say that companies are allowed to collect data and turn it into something of value (ibid). This is the logic that industries are using to justify appropriation; user data is not owned by the user itself but is ‘there for the taking’ by companies to process into something of value for their users and others. But in order to claim this, the conditions for such a biopolitical public domain have to be created. The final section of this article looks to understand and explain the implications of this shift against a more holistic assessment of extractivism by publishers.

The turn to data extraction by academic publishers

Through its attempts to keep users interacting with publisher-provided services, GetFTR individuates users in a way that places researchers, rather than published content, as the focal point of scholarly communication. This represents a broader ontological shift in academic publishing towards the user and their research behaviours. Publishers now seek user

integration and interaction with all elements of the end-to-end publishing workflow from personalised research recommendations to evaluative metrics and beyond. Through a process of vertical integration of products and services, publishing companies are able to track researcher data from pre-submission to publication and beyond (Posada & Chen, 2018).

For example, having recently rebranded itself from publisher to ‘information analytics’ business, Elsevier is at the forefront of the shift towards user data. The CEO of Elsevier, Kumsal Bayazit, describes its strategy to ‘build data, acquire it, license it, aggregate it, and then build contributed databases where their customers can contribute’ (van Loon, n.d.). Alongside its extensive journal portfolio, Elsevier has access to many millions of citation records through its Scopus databases, which Bayazit describes as ‘a rich data set for us to mine’ (ibid). In addition to this, Elsevier has been strategically acquiring startup businesses across the entire research and publication lifecycle, including altmetrics company Plum Analytics, citation manager Mendeley, preprint server Social Science Research Network and university repository platform Bepress, among many others (Posada & Chen, 2018). These acquisitions are not valued for the content they create but the data analytics they generate through user interactions across Elsevier’s products and services.

A recent landscape report from the Scholarly Publishing and Academic Resources Coalition (SPARC) illustrates the shift to data analytics by the publishing industry. Rather than pursuing acquisitions of publications, Elsevier has instead broadened its products to reach new markets, explicitly indicating its ‘intent to increasingly serve university administrations, funding bodies, and governments with tools aimed at estimating and improving the productivity of research and optimizing funding decisions’ (Aspesi et al., 2019, p. 14). This means that Elsevier, and other big publishers, are looking to broaden their customer base

beyond the library and into other areas of the university and higher education, especially to those interested in measuring performance or understanding researcher behaviours (Lamdan, 2019). As Penny Andrews writes, ‘the real threat may not be publisher control of research outputs, but of workflows and data flows, which are far more difficult for either researchers or librarians to ethically disrupt’ (Andrews, 2018, p. 208). This practice is all the more sinister when libraries are bypassed entirely through national deals for open access, such as in the Netherlands where Elsevier receives access to user data in exchange for providing open access to its content (Elsevier, 2019).

The data-driven business models of publishers are therefore more holistic in how they relate to the entire research lifecycle and more expansive in terms of their potential customer base. Alongside universities, funders and governments, publishers are also able to share data across all their parent companies’ holdings. For example, Sarah Lamdan explores how Elsevier’s parent company RELX is seeking to amass and process huge caches of data across their company, including its ‘legal solutions’ company LexisNexis that is ‘selling the surveillance products that help law enforcement track, detain, and charge people with crimes’ (Lamdan, 2019). Part of this process, Lamdan argues, entails surveillance of library patrons, many of whom would be unaware that their data are being packaged and sold to a variety of companies with questionable ethical practices.

So, the ultimate aim of GetFTR is part of a larger project to stabilise users by individuating them such that they can be traced across the products and services of commercial publishers looking to monetise the data generated. Roger Schonfeld has explored the possibility that the move towards seamlessness may in fact be an attempt to turn publisher access portals into ‘supercontinents’ whereby users may access ‘a truly central platform for discovery, access,

and use' from across all publishers (Schonfeld, 2018). Schonfeld writes: 'access is difficult, since authorization must take place separately for each platform. The personalization that researchers crave to improve their productivity is challenging, since user accounts are siloed on a platform basis. Band-aid solutions like RA21 will do nothing more than attempt to build bridges across archipelagos, rather than solving the underlying fragmentation problem' (Schonfeld, 2018, p. unpaginated). For Schonfeld, one solution would be to allow users to access all content through one centralised account (a 'supercontinent' rather than a series of interconnected 'archipelagos'), rather than maintaining individual accounts connected through a service like GetFTR.

When taking into account publishers' desire to create a biopolitical public domain of accessible and monetizable data, the supercontinent metaphor seems a likely prospect for the future of publishing. In this instance, the user itself would be completely trackable across all participating publisher products and services, allowing full appropriation of their interactions for monetary gain. The supercontinent metaphor is also helpful when read against what Couldry and Mejias describe as the turn towards 'data colonialism' within digital capitalism:

'[D]ata colonialism is our term for the extension of a global process of extraction that started under colonialism and continued through industrial capitalism, culminating in today's new form: instead of natural resources and labor, what is now being appropriated is human life through its conversion into data. The result degrades life, first by exposing it continuously to monitoring and surveillance (through which data is extracted) and second by thus making human life a direct input to capitalist production. Data colonialism is, in other words, an emerging order for appropriating

and extracting social resources for profit through data, practiced via data relations’
(Couldry & Mejias, 2019, p. xix).

Data colonialism is, for Couldry and Mejias, neoliberalism taken to its logical conclusion: the opening up (‘literally annexing’) of all areas of life to monetisation as data (Couldry & Mejias, 2019, p. 33). It is an exercise of power that makes the individual wholly available for capital, resulting in the ‘degradation’ of life through continual surveillance.

Although Couldry and Mejias are keen to emphasise that ‘colonialism’ is not deployed as a metaphor (2019, p. xi), the term fits neatly for my purposes with the idea of the future of research access as a supercontinent, or terrain to be colonised by the publishing oligopoly. By making it easier for researchers to access content, and thus control their access pathways, users become raw materials that can be extracted, processed and used in a variety of ways – either to tailor services or to sell to external customers. Publishers seek end-to-end scholarly communication ecosystems in order to colonise the data of researchers by keeping them interacting on these platforms to the greatest extent possible. Data colonialism is therefore the logical extension of GetFTR.

GetFTR and researcher quantification in the university

It is important to note that universities are complicit in the process of data colonialism; publishers are selling a product that universities consider highly valuable. The higher education sector is one of the primary customers of the data analytics produced by publishers and analytics providers. In a landscape report on the usage of metrics in international research assessment, Wilsdon et al. identified a range of ways in which quantitative metrics are sought

after for research assessment, including: ‘demands by policymakers for more strategic intelligence on research quality and impact; the need for institutions to manage and develop their strategies for research; competition within and between institutions for prestige, students, staff and resources; and increases in the availability of real-time ‘big data’ on research uptake, and the capacity of tools for analysing them’ (Wilsdon et al., 2015, p. vii). The authors highlight how publishing and analytics companies have responded to these needs through altmetric services, CRIS (Current Research Information Systems), research data management platforms and other performance tracking software (Wilsdon et al., 2015, p. 39).

The individuation of users through interventions like GetFTR is illustrative of the turn towards quantification in higher education. As universities increasingly rely on empty notions of ‘excellence’ to which academics should adhere (Moore et al., 2017; Readings, 1996), there is greater measurement and concentration of practices that purport to reflect this excellence. Alongside traditional indicators of quality such as journal prestige and Impact Factor, weaknesses notwithstanding, researchers are also now assessed by a broader mixture of quantitative indicators such as article-level metrics, funding awards and citation counts. Universities want to be able to track this in order to understand who is adequately performing and, crucially, who is not. The data-driven insights sold by publishers may allow universities to predict career trajectories and understand which areas of scholarship are more marketable to prospective students (Aspesi et al., 2019, p. 16).

Much like with GetFTR, the move towards researcher quantification and tracking is collaborative between major publishing and technology companies, rather than purely based on competition, further revealing the desire by the publishing industry to create the conditions for a biopolitical public domain of available and extractable data. The ‘Open

Researcher Contributor Identification Initiative' (ORCID) is another example of this collaborative approach to creating a biopolitical public domain. A not-for-profit multi-stakeholder collaboration, ORCID is funded by large commercial publishing houses, including the publishing oligopoly discussed here, alongside many other not-for-profit institutions, universities and research institutes. The initiative started out as a simple and seemingly benign project to disambiguate researchers by assigning them a unique alphanumeric code associated to associate with their publications. This allows researchers to still obtain credit for their publications when, for example, they move institution, change their name, or share a name with another researcher: the alphanumeric code follows acts as the unique identifier in each case.

Yet ORCID is now integral to the performance tracking of universities across the world. Many journals have made ORCID a mandatory feature of their submission system infrastructures, while entire national higher education systems are requiring their academics to sign up for the service (Carter & Blanford, 2017; Scott & Thomas, 2018). What began as an opt-in tool to disambiguate authorship is now an often-mandatory component of higher education's disciplinary mechanisms. Like GetFTR, this is achieved through individuating researchers in a way that allows them to be trackable by universities as single, quantifiable units. Universities, publishers and other actors in higher education are all working towards the same goal, but often for different reasons. For universities, quantification affords the potential to understand researcher performance, while the publishing oligopoly seeks consolidation of their market power. Researchers may even desire many of the features that promote individuation, such as increased interoperability of services, ease of research access, and the ability to track and present their research outputs. But this benefits will ultimately come at the price of greater surveillance and data extraction, or what Shoshana Zuboff terms

a ‘twenty-first century Faustian compact’ due to the impossibility of resisting these services (Zuboff, 2019, p. 11).

Concluding remarks

Much like the single-sign-on universes of Google, Facebook and Amazon, commercial scholarly communication is rapidly moving to an era in which the individual user, rather than the published content, is positioned as the focal point of the publishing industry. The seemingly small infrastructural intervention of *Get Full Text Research* is indicative of a broader trend to individuate researchers in order to keep them within the walled gardens of the academic publishing oligopoly, thus preventing them from accessing research in other ways (illicit and otherwise). Ostensibly, GetFTR is promoted to help researchers easily access content without needing to navigate cumbersome authentication systems, yet this comes with the added benefit to publishers of keeping researchers continually interacting with their products and services. With some commentators predicting a move towards a single platform (‘supercontinent’) for all research access, the five largest publishers – theorised here as the publishing oligopoly – stand to benefit considerably from this new arrangement, especially as they increasingly move towards the provision of data analytics alongside published content.

I have theorised this move as indicative of the desire by the publishing oligopoly to create a biopolitical public domain of user data that is extractable through practices of data colonialism (Cohen, 2018; Couldry & Mejias, 2019). Importantly, this newly extractable resource is created collaboratively with other publishers, not just through GetFTR but also ORCID and other initiatives too. Universities and governments are encouraging this shift through their desire for analytics based on researcher outputs and performance, reflective of a

hypercompetitive environment where quantitative measures are required to understand researcher performance. The result of this practice of data colonialism, as Couldry and Mejias write, ‘degrades life, first by exposing it continuously to monitoring and surveillance (through which data is extracted) and second by thus making human life a direct input to capitalist production’ (Couldry & Mejias, 2019, p. xix). This is the direction in which publishing is heading.

Higher education is not unique in the move to data extraction and surveillance, nor have I explored the full implications of it here, and countering the threat will provide challenging as in other industries. Yet academic publishing is unique in how much of the editorial labour and content is provided by academics for free, traditionally conceived as part of the academic job, and this offers the opportunity for academics to demand certain conditions of the publishing industry. The ability of publishers to design systems to promote data colonialism is largely down to the fact that universities have ceded control of scholarly communication infrastructures to private, for-profit companies, through a process of marketisation that was to hasten in the 1980s and culminate in the system that exists today. Advocates of more ethical publishing systems initially conceived of open access as a way to hold publishers accountable, though in practice open access has allowed commercial publishers to increase their stranglehold through article-processing charges and has even provided the conditions for data extraction (as explored here).

Instead, the turn to data extraction requires an intervention into the free market practices of academic publishers, including through systems of governance and data trusts and privacy policies managed by and with academic communities themselves. This means that academics should be able to collectively demand conditions of publishers around how data is collected

and used, and to whom it is sold. In a roadmap document on combatting the threat of privately-owned academic infrastructure, the Scholarly Publishing and Academic Resources Coalition (SPARC) divide these interventions into ‘risk management’, ‘strategic choices’ and ‘community actions’, a holistic approach to countering the threat (SPARC, 2019). There are also many initiatives underway to build open-source publishing infrastructures that do not permit this kind of data extraction, by organisations such as Invest in Open Infrastructure, the Community-led Open Publishing Infrastructures for Monographs (COPIM) and the cooperatively-managed Latin American technology consortium Amelica. Many of these operate on a fraction of the budgets of large commercial publishers and may struggle to compete for support without broader academic engagement.

But ultimately, the push to individuate and track researchers is not purely an external threat from commercial publishers. As I have shown, it is a demand of the higher education sector more generally and will therefore require a political solution. Universities want to track researchers because they are primarily acting as businesses looking for an economically efficient way of making decisions and measuring staff performance. Academics are forced to rehearse cultures of individualism due to the need to participate in a hypercompetitive academic job market that works against collective action by promoting self-interest above all. This is to say that universities are culpable for the turn to extractivism: academic publishers are merely serving this culture and promoting new and invasive ways of monetisation through data extractivism.

References

- Andrews, P. C. S. (2018). “Putting It Together, That’s What Counts”: Data Foam, a Snowball and Researcher Evaluation. In P. V. Moore, M. Upchurch, & X. Whittaker (Eds.), *Humans and Machines at Work* (pp. 203–229). Springer International Publishing. https://doi.org/10.1007/978-3-319-58232-0_9
- Arsenault, J., Dresselhaus, A., Tokoro, S., & Twardowski, K. (2020). The Authentication Landscape in 2019: One Does Not Simply Walk into Order. *The Serials Librarian*, 1–5. <https://doi.org/10.1080/0361526X.2020.1728733>
- Aspesi, C., Allen, N., Crow, R., Daugherty, S., Joseph, H., McArthur, J., & Shockey, N. (2019). *SPARC 2019 Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions*. SPARC. <https://sparcopen.org/our-work/landscape-analysis/>
- Bohannon, J. (2016). Who’s downloading pirated papers? Everyone. *Science / AAAS*. <https://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>
- Brainard, J. (2019, November 15). Move by journals to ‘seamless’ off-campus access raises privacy concerns. *Science / AAAS*. <https://www.sciencemag.org/news/2019/11/move-journals-seamless-campus-access-raises-privacy-concerns>
- Buranyi, S. (2017, June 27). Is the staggeringly profitable business of scientific publishing bad for science? *The Guardian*. <http://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science>
- Carter, C. B., & Blanford, C. F. (2017). All authors must now supply ORCID identifiers. *Journal of Materials Science*, 52(11), 6147–6149. <https://doi.org/10.1007/s10853-017-0919-7>

- Cohen, J. E. (2018). The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy. *Philosophy & Technology*, 31(2), 213–233.
<https://doi.org/10.1007/s13347-017-0258-2>
- Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. /z-wcorg/.
- Elsevier. (2019). *Dutch research institutions and Elsevier reach framework agreement*.
<https://www.elsevier.com/about/press-releases/corporate/dutch-research-institutions-and-elsevier-reach-framework-agreement>
- Eve, M. P. (2014). *Open access and the humanities: Contexts, controversies and the future*. Cambridge University Press.
- Fried Foster, N. (2014). *Information Literacy and Research Practices*. Ithaka S + R.
https://sr.ithaka.org/wp-content/uploads/2014/11/SR_Briefing_Information_Literacy_Research_Practices_20141113.pdf
- Get FTR. (n.d.). *Introducing Get Full Text Research*. GetFTR. Retrieved 22 February 2020, from <https://www.getfulltextresearch.com/introducing-getftr/>
- Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., Greshake Tzovaras, B., & Greene, C. S. (2018). Sci-Hub provides access to nearly all scholarly literature. *ELife*, 7, e32822. <https://doi.org/10.7554/eLife.32822>
- Hinchliffe, L. (2019, December 10). Why are Librarians Concerned about GetFTR? *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2019/12/10/why-are-librarians-concerned-about-getftr/>
- Jamali, H. R. (2017). Copyright compliance and infringement in ResearchGate full-text journal articles. *Scientometrics*, 112(1), 241–254. <https://doi.org/10.1007/s11192-017-2291-4>

- Jamali, H. R., & Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics*, 105(3), 1635–1651.
<https://doi.org/10.1007/s11192-015-1642-2>
- Lamdan, S. (2019). Librarianship at the Crossroads of ICE Surveillance. *In the Library with the Lead Pipe*. /2019/ice-surveillance/
- Larivière, V., Haustein, S., & Mongeon, P. (2015a). *Big Publishers, Bigger Profits: How the Scholarly Community Lost the Control of its Journals*. 2, 9.
- Larivière, V., Haustein, S., & Mongeon, P. (2015b). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE*, 10(6), e0127502.
<https://doi.org/10.1371/journal.pone.0127502>
- McKenzie, L. (2019, March 1). University of California cancels deal with Elsevier after months of negotiations. *Inside Higher Education*.
<https://www.insidehighered.com/news/2019/03/01/university-california-cancels-deal-elsevier-after-months-negotiations>
- Moore, S., Neylon, C., Paul Eve, M., Paul O'Donnell, D., & Pattinson, D. (2017). "Excellence R Us": University research and the fetishisation of excellence. *Palgrave Communications*, 3, 16105.
- Morozov, E. (2018, August 19). There is a leftwing way to challenge big tech for our data. Here it is | Evgeny Morozov. *The Guardian*.
<https://www.theguardian.com/commentisfree/2018/aug/19/there-is-a-leftwing-way-to-challenge-big-data-here-it-is>
- Mulvany, I. (2019, December 11). Thoughts on GetFTR. *ScholCommsProd*. http://scholarly-comms-product-blog.com/2019/12/11/thoughts_on_getftr_/
- Murray, P. (2019, December 28). What is known about GetFTR at the end of 2019. *Disruptive Library Technology Jester*. <https://dltj.org/article/getftr-update/>

NISO. (2019). *Recommended Practices for Improved Access to Institutionally-Provided Information Resources* (p. 49).

https://groups.niso.org/apps/group_public/download.php/21892/NISO_RP-27-2019_RA21_Identity_Discovery_and_Persistence.pdf

Open Access Button. (2019, December 5). Get To Fulltext Ourselves, not GetFTR. *Medium*.

<https://blog.openaccessbutton.org/get-to-fulltext-ourselves-not-getftr-e952e798564b>

Piwowar, H., Priem, J., & Orr, R. (2019). *The Future of OA: A large-scale analysis projecting Open Access publication and readership* [Preprint]. Scientific Communication and Education. <https://doi.org/10.1101/795310>

Posada, A., & Chen, G. (2018, June 15). Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers. *22nd International Conference on Electronic Publishing*. 22nd International Conference on Electronic Publishing. <https://doi.org/10.4000/proceedings.elpub.2018.30>

Readings, B. (1996). *The University in Ruins*. Harvard University Press.

ResearchGate. (n.d.). *ResearchGate / Find and share research*. ResearchGate. Retrieved 4 March 2020, from <https://www.researchgate.net/>

Sanchez, H., Murray, J., & Sanchez, D. (2016). Putting Your Passwords on Self Destruct Mode. *Soups 2016: Twelfth Symposium on Usable Privacy and Security*, 5. <https://www.usenix.org/conference/soups2016/workshop-program/wsf/presentation/sanchez>

Schonfeld, R. (2018, May 3). One platform to rule them all? *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2018/05/03/supercontinent-scholarly-publishing/>

- Schonfeld, R. (2019a). Publishers Announce a Major New Service to Plug Leakage. *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2019/12/03/publishers-announce-plug-leakage/>
- Schonfeld, R. (2019b, March 7). Is the Value of the Big Deal in Decline? *The Scholarly Kitchen*. <https://scholarlykitchen.sspnet.org/2019/03/07/value-big-deal-leakage/>
- Scott, A., & Thomas, H. (2018). Chapter 15—The Implementation of ORCID by All Major Research Institutions Across New Zealand. In J. Atkinson (Ed.), *Collaboration and the Academic Library* (pp. 165–172). Chandos Publishing.
<https://doi.org/10.1016/B978-0-08-102084-5.00015-8>
- SPARC. (n.d.). *Big Deal Knowledge Base*. SPARC. Retrieved 4 March 2020, from <https://sparcopen.org/our-work/big-deal-knowledge-base/>
- SPARC. (2019). *SPARC Roadmap for Action.pdf*. SPARC. <https://sparcopen.org/our-work/roadmap-for-action/>
- Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, 43(3).
<https://doi.org/10.1177/00027649921955326>
- van Loon, R. (n.d.). *RELX Group: A transformation story*. RELX. Retrieved 28 March 2020, from <https://www.relx.com/our-business/our-stories/transformation-to-analytics>
- Van Noorden, R. (2017). Publishers threaten to remove millions of papers from ResearchGate. *Nature News*. <https://doi.org/10.1038/nature.2017.22793>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <http://doi.org/10.13140/RG.2.1.4929.1363>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. Public Affairs; /z-wcorg/.

